DOCUMENT RESUME

ED 286 376 FL 016 926

AUTHOR Ittzes, Kata
TITLE A Closer Look at Cloze.
PUB DATE Apr 87
NOTE 17p.; Paper presented at the Annual Meeting of the International Association of Teachers of English as a Foreign Language (21st, Westende, Belgium, April 12-14, 1987).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Classroom Research; *Cloze Procedure; Comparative Analysis; *English (Second Language); Foreign Countries; Hungarian; *Language Tests; Language Usage; *Scoring; Second Language Instruction; Teacher Attitudes; Test Format; *Test Use; Uncommonly Taught Languages
IDENTIFIERS Hungary

ABSTRACT

A Hungarian teacher of English as a foreign language (EFL) chronicles her research on the use of the cloze procedure in her native country's EFL classes. The research began as an effort to convince other teachers of the test format's usefulness, and ultimately led to further study of the factors affecting responses and scoring and the comparability of responses on English and Hungarian language cloze tests. Some data from the research are presented. A major conclusion arrived at is that this type of research is not only appropriate but manageable for classroom teachers and can contribute to the fields of both linguistics and language teaching. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED286376

A CLOSER LOOK AT CLOZE

Kata Ittzes

Westende, Belgium

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

K. Ittzes

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

FL016926

Paper presented at the Annual Meeting of the International Association of Teachers of English as a Foreign Language (21st, Westende, Belgium, April 12-14, 1987).

BEST COPY AVAILABLE

Kata Ittzés

IATEFL 1987
Westende, Belgium

A Closer Look at Cloze

In Hungary, a few years ago cloze tests were and, in most places, still are a novelty. So before I first introduced cloze tests in a county English competition, I tried out a Hungarian cloze with teachers and advisers in Hungarian literature and language during a 3-day workshop on testing in 1983. My primary aim with the test was to point out to the educational authorities that this type of test cannot yield very high scores so that they should not blame the competitors, the test, or the level of ELT in our county when they saw the test scores in the county competition. I used material at hand, the beginning of a short story by a contemporary Hungarian writer found in a secondary school textbook. The first reaction usually was, "Oh, I've known that story and have been teaching it for years so my solution wouldn't be of any use to you." The scores of the 11 testees ranged from 30 to 43 on the 50-gap test. The body of data, which was later considered to be a part of a pilot test, revealed a few problems and hypothetic findings which still have to be proved on more data.

1. In the literary text, people who had actually done creative writing did better than people who had been teaching the story in question. This agrees with Oller's view: "Supplying the missing word is a constructive process similar to the way the writer put the text together." /Oller 1979:343/. In addition to the teachers and advisers I asked several members of my family to do the test - people who had not known the story but were well read and who, because of their jobs, had to write a lot /editorial work, reviews/. It would be an interesting point of further research to examine whether people who are engaged in writing do better on cloze tests than others. In a cloze test the testee has to be creative, has to "get inside the language" /Radice

"dressed" while none of the "writers" did. At one place a sergeant says something like "It's none of your business" to the prisoner. After the quote almost everybody wrote "said the sergeant". The original was something like "grumbled" or "mumbled". If we only accept "said" as the most usual variant, in addition to the original, we exclude lots of colourful, expressive synonyms which may come up in a large body of test data. So what should the criterions of acceptability be? /In the above mentioned gap there were three correct solutions, one of which came from the pupil with the worst ENG test. This item yielded one of the two negative discriminations of the HUN test./

The findings of the pilot test led me to decide that when there are several examiners, when the test is administered in a number of different schools, only the exact scoring procedure will be used. A comparison of CE and CA in the pilot test showed that only two out of eleven differed on the two rank-lists. Since it was hypothesized and later proved that there would be far fewer acceptable solutions in the ENG cloze than in the HUN, we could say that CE in a FL gives a fairly accurate ranking. Research in the U.S. has shown a very high correlation between CE and CA /e.g. Hinofotis 1980, Oller 1979a/.

For two years, in 1984 and 1985 CE had been used in the county English competition alongside with a discrete-point multiple choice /MCH/ test. Practice material in cloze was sent out to schools and reasons for CE were explained. Although cloze results showed high correlations with MCH, in these first years there was some reluctance on the part of a few teachers to accept cloze as a valid measure of their pupils' proficiency. This reluctance was most evident in the case of teachers with a keen interest in formal grammar and structural exercises. I received letters pressing me to accept certain solutions teachers found correct. E.g. when a bundle of newspapers was wrapped

1978:203/. The intelligent testee often puts in a better word than the original /Porter 1981, Oller 1979:343/.

2. In the pilot data primary school advisers scored remarkably worse than secondary school teachers and advisers. The primary school adviser with the highest score admitted that, although he wasn't teaching the story in question in primary school, the writer was his favourite.

3. Another finding was that, contrary to most American and British statements, it was not easy to agree in advance upon acceptable solutions. /Research backs this finding. Hughes /1981/ had to discard more than half of his items because native speakers were not in agreement on them. Gefen /1979/ also found that native judges disagreed./ Having done the test, I asked the advisers to discuss their solutions. Sometimes synonyms resulted in a debate - what kind of gait can a captain of the Hungarian armed forces in the 1950's walk with in a small office room in a prison? The original was "slow", synonyms for "small" are all right but what about "huge" and its synonyms? In the final data there were 26 variants for this word, two of which only were definitely unacceptable.

4. I had to face another question. Is it possible that a word everybody is using is unacceptable? What are the norms of acceptance? Let me still refer to the Hungarian text partly because I move safer in my native language, partly because the ENG data showed very few acceptable synonyms. In the HUN test the prisoner, who is to be freed, was led into an office where there had already been 8-10 people "dressed" in civilian clothes according to 139 of the 146 testees. These prisoners in the original were "dressed back into" /literal translation, one word only/ or "changed". Prisoners are not normally "dressed" in civilian clothes. It is interesting to note that all the teachers accepted

Royal Parks provide a welcome rest for the visitor." Why not "visitors", "tourists", asked several teachers.

Last year, in January 1986, the pupils were asked to fill in a HUN cloze test besides the two other tests. I think this test made teachers realize the problems of cloze scoring and the value of the English cloze. When they saw that native speakers of Hungarian were not at all "perfect" in a test given in Hungarian and when they could also see that pupils who were generally thought to be doing better scored higher on HUN cloze than others who were considered low achievers, teachers started to believe in cloze testing. It is interesting to see now that I have evaluated a part of the data that the HUN cloze discriminates far worse than the ENG, still, the "feel" of the native language was needed to make teachers and pupils accept this form of testing. The Appendix contains the α indices of discrimination for the three tests. You can also see the data about the tests and the testees there.

When the tests had been written, I was left with some hypotheses, two shopping bags full of test papers, long nights and free weekends, a search for an adequate computer programme /which, when I thought I had found it, proved unsuitable/ and a pocket calculator. My research then is far from finished, most of my hypotheses still need proving or disproving but - and this can be taken as a positive element - most of the existing evaluation has been done with tools accessible to any teacher in any school. So, after outlining my hypotheses, I would like to show you how I went about evaluating my data and what linguistic problems I have come across in this test data.

Some of the hypotheses I want to examine are as follows:

1. Those who do poorly on the HUN cloze, do not do well on the ENG cloze;
2. Those who do well on the HUN cloze, do not necessarily do well on

the ENG cloze /there are too many factors involved: interest in English, quality of teaching, learning time, etc./;

3. The ENG cloze and the MCH will correlate better than the HUN cloze and either of the above - see 2;
4. The HUN cloze - ENG cloze correlations will be higher for more advanced learners than for beginners. If this is proved to be true we can conclude that beginners will develop in time in a similar way. If the correlations for more advanced learners are considerably high, HUN cloze can form a part of examinations to decide who should take part in specialized FL tuition.

Though I was unable to evaluate my test data properly, the table in the Appendix shows if not very scientific but, at least for me, impressively symmetric figures. I could not attempt more with my crude measuring instruments than to compare the best and the worst 20 tests on the HUN and the ENG cloze. I think these data show that there must be some correlation between the two tests.

The relatively low exact solution of the HUN cloze /63.3/ where there were no items with 100 p.c. and two with 0 /whereas in the ENG test there were three items with 100 p.c. solution and only one with 0/ and where 22 pupils did better on the ENG test than on the HUN led me to reconsider the choice of my HUN text. Although it was contemporary prose with lots of dialogue, it was still too literary, subtle and culturally-historically loaded compared with the Grimm tale. I was anxious to choose an ENG text which was suitable for the linguistic level of most second year pupils and also one which was on a topic not completely unknown to them. The ENG test was probably a bit too easy and the HUN test proved to be difficult for the testees.

Returning to my other hypotheses and later findings, I was sure that the CA would yield much higher scores in the HUN than in the

ENG cloze where CA does not substantially improve the scores. I also predicted that there would be a great number of non-responses in the ENG while very few in the HUN and that the range of answers in certain ENG and HUN items would be due to different factors. In the case of the HUN text the number of acceptable solutions can be very high, in the ENG cloze the high range does not equal a great number of acceptable solutions. E.g. the most varied item for gap 1 /loved/ yielded 28 kinds of answers only two of which /liked, preferred/ could be accepted. The highest ranges in the HUN cloze showed synonyms, in the ENG they indicated confusion. In the ENG test there was a much higher number of wrong word class, wrong form or wrong word unacceptability than in the HUN. It is a future task to ask native judges to decide on acceptability. This is just being done for the Hungarian test.

The two tests were compared from the point of view of item facility values /FV/, discrimination indices, and the number of non-responses for different parts of speech. No two tests would show exactly the same order of difficulty for various parts of speech but there are some general tendencies. The Lapkin and Swain investigation of the Ontario Institute /Lapkin and Swain 1977/ of immersion school results show almost the same order in French as my research in English.

When looking at the discriminatory power of items I found that items with the highest discrimination index are unambiguous and have only one possible solution or can be predicted from the context. It is a surprising finding that there were no HUN items discriminating better than .50 /there were only two above .40/ while there were 15 items on the ENG test with higher than .50 discrimination. Items with low discrimination are either too easy or too difficult. A strong correlation was found between the number of blanks and the FV of a certain item. Blanks are characteristic of certain

parts of speech as well as of individual items. There were very few non-responses where there was a high number of correct answers and the opposite was also true. There was one exception: almost everybody attempted to fill in gap 44 /outside, 3 blanks/ and only one person found the exact solution. In general, in items with fewer than 10 blanks the result was 80-100p.c., in items with more than 50 blanks the testees scored 3-35 p.c.

So far I have given you lots of figures but can we learn anything from these figures and can we learn anything else from cloze tests than figures? I am convinced that even the numbers reveal facts about the language and the learning of a language but there are other factors as well. Whatever text one chooses one can find items to analyze which are of more general interest than the given text itself. Porter says that cloze tests will also sample items which we know only unconsciously /Porter 1981/. Analyzing a cloze text one always comes across such items. It is interesting to look out for common usage with native speakers. How many would use "forest", and how many would say "wood" or "woods" in gap 29? What is the proportion of "anybody"-"anyone" in gap 12? Another point of investigation can be the same word in a different context. What makes "you" different in gaps 26 and 27? There are more than twice as many good solutions for No. 27 than for 26. /"You look in front of <u>you</u> when you're walking, just as if <u>you</u> were going to school."/ Between the two, "you" appears once and "were" probably also helps after the second "you".

There are interesting misunderstandings carried on. Before gap 3, some pupils thought that "cap" meant "cup" so they filled gap 3 with "wine" /red wine, to be more precise/, "glass" and "bottle". This misunderstanding was carried as far as gap 5 where "drink" cap occurred. A similar solution, characteristic of beginners who

do not see more than a few words before or after a gap and cannot make use of discourse clues, appears in gaps 26-27-29. Here "school", which appears in the text between gaps 27-28 influences 26 /school/, 27 /children/ and 29 /school, this time in six cases/. Others, thinking in a larger context, picked the word "flowers" /after gap 22/. They filled gap 29 with "flowers"/"vase"/"garden".

Set patterns influence No. 47. After "just", both "minute" and "minutes" appear without any contextual reference. This leads us to a problem with articles as well. In the HUN test half of the unacceptable solutions concern articles. An extra or a missing article are very easy to overlook. If the testee remembers the phrase "just a minute", she overlooks the lack of an "a" when using "just minute". A similar place is "all the flowers" in gap 31, where 11 testees wrote "of" /all of the.../. In gap 11, many use an adjective instead of the article in front of "path" /green, right, straight, wood, etc./. The same happens in gap 37. Instead of "into the wood" the testees go into "dark" or "forest" wood. Indefinite articles never replace the definite article but the opposite often happens. In No. 19. /"For a little while"/ 34 pupils write "the", in 33 /"a bunch of flowers"/ 14 write "the, 12 write "some". Indefinite articles are much worse, more difficult. The average FV of definite articles is 76.5 while that of indefinite articles is 53.7.

I do not want to quote more figures. What I have wanted to pass on more than my actual findings is the belief that cloze can offer insights into and revelations about language. It is something like taking a machine apart to see how it works. I am also convinced that this kind of research is most suitable for teachers. Linguists rarely go into the classroom where we teachers all have a wealth of data which is there for the asking. By collecting and evaluating this data we can help linguistics in our modest was to deal with

issues closer to our world, the world of language teaching. John Oller has helped me a lot and can help other teachers as well by making us believe that we may have something to offer for linguistics. "Surely the findings of classroom teachers /especially language teachers or teachers who have learned the importance of language to all aspects of an educational curriculum/ are as important to the theories of language as the theories themselves are to what happens in the classroom. Unfortunately, the direction of influence has been much too one-sided." /Oller 1979:151/ It would be too much to ask ourselves to balance this difference. Our main task is teaching, not doing research. But there is one more advantage in classroom research: real learning is always some kind of research. By doing more than the daily routine of teaching we may retain in ourselves the quality of the learner.

References

GEFEN,R.1979. An experiment with Cloze Testing. English Language Teaching Journal 33,2,122-126.

HINOFOTIS,F.B.1980. Cloze as an Alternative Method of ESL Placement and Proficiency Testing. In Oller and Perkins eds. Research in Language Testing. Newbury House, Rowley,Mass.,121-128.

HUGHES,A.1981. Cloze as a Measure of Oral Ability. ELTJ 35,2,161-168.

OLLER,J.W.Jr.1979. Language Tests at School. Longman.

OLLER,J.W.Jr.1979a. Intelligence, Achievement, Competency or Language? In Forio, Perkins and Schachter eds. On TESOL '79. The Learner in Focus. TESOL, Washington D.C.,40-54.

PORTER,D.1981. Prose Cloze, Oral Cloze and Conversational Cloze. Unpublished lecture. Reading.

RADICE,F.W.1978. Using the Cloze Procedure as a Teaching Technique. ELTJ 32,3,201-204.

CLOZE/86 ~~Kata Ittzés, Hungary~~

N=146 secondary school pupils aged 16. 95 beginners /2nd year of English/, +1-3 yrs:32, +6 yrs:19

3 tests: HUN/garian cloze/ ENG/lish cloze/ MCH /discrete point multiple choice/

Max. score: 50 points for each test

Average scores for the whole population:	HUN	ENG	MCH
	31.68	23.2	28.38
α indices:	.630	.886	.855

Average scores of

	HUN best 20	HUN worst 20	ENG best 20	ENG worst 20
HUN	37.75 /+6.1/	24.84 /-6.8/	33.8/+2.1/	29.5 /-2.1/
ENG	26.3 /+3.1/	20.07 /-3.1/	34.8 /+11.6/	11.9 /-11.3/
MCH	29.45 /+1.1/	23.34 /-5.0/	38.75 /+10.4/	19.65 /-9.7/

Distribution of responses

HUN				ENG			
4619	63.3		Correct /exact/	3396	46.5		
95.6 2357		32.2	Acceptable	57.0 762			10.4
210	35.1	2.9	Unacceptable	1804	35.2		24.7
114	1.6		Non-response	1338	18.3		

Average FV's, discrimination, number of non-responses for different parts of speech

HUN				ENG			
Conjunctions	94.2	.19	0	Preposition	89.0	.35	4
Articles	88.8	.24	2	Articles	68.9	.43	21
Pronouns	66.3	.15	0	Pronouns	60.4	.38	15
Nouns	62.1	.24	3.6	Nouns	57.7	.10	18
Verbs	50.2	.18	1.2	Conjunctions	53.9	.43	33.2
Adverbs	46.1	.13	3	Adverbs	18.9	.29	39.3
Adjectives	27.6	.14	3	Adjectives	6.6	.32	25.7

Range of scores: HUN 41 - 18 /best 20: 41-36, worst 20: 27-18/
ENG 42 - 7 /best 20: 42-34, worst 20: 14-7/
NCH 45 - 9

APPENDIX B

LITTLE RED RIDING HOOD

Once upon a time there was a sweet little girl who everybody loved. But most of all she was /1/_____ by her grandmother. One day she /2/_____ her a cap made of red /3/_____ and because it suited her so /4/_____, and because it was the girl's /5/_____ cap, she was called Little Red /6/_____ Hood.

One day her mother said /7/_____ her, "Come here, Little Red Riding /8/_____, take this piece of cake and /9/_____ bottle of wine to your grandmother." /10/_____ mother told her to stay on /11/_____ path and not to talk to /12/_____ because the woods were full of /13/_____ animals.

The wolf thought to himself, "/14/_____ young, tender thing will make a /15/_____ meal. She will taste even better /16/_____ the old woman. But I must /17/_____ very cunning, so that I can /18/_____ hold of both of them." For /19/_____ little while he walked next to /20/_____ Red Riding Hood and then he /21/_____, "Little Red Riding Hood, look at /22/_____ beautiful flowers here. Why don't you /23/_____ around? I don't think you even /24/_____ the beautiful songs the birds are /25/_____! You look straight in front of /26/_____ when you're walking, just as if /27/_____ were going to school. Don't you /28/_____ how pleasant it is in the /29/_____?"

Little Red Riding Hood looked round /30/_____ when she saw how beautiful all /31/_____ flowers were, she thought, "Grandmother would /32/_____ very happy if I took her /33/_____ bunch of fresh flowers. It is /34/_____ early in the day that I /35/_____ have plenty of time to get /36/_____."

She went off the path into /37/_____ wood and looked for flowers. Whenever /38/_____ picked one, she would spot a /39/_____ beautiful specimen further away, and she /40/_____ after it and

so went deeper /41/_____ deeper into the forest.

The wolf, /42/_____, went straight to the grandmother's house /43/_____ knocked on the door. "Who is /44/_____?", she asked. "Little Red Riding Hood. /45/_____ am bringing you cake and wine. /46/_____ the door," said the wolf. "Just /47/_____ down the handle," the grandmother replied, "/48/_____ am too weak and cannot get /49/_____ from my bed." The wolf pushed /50/_____ door handle down and opened the door. Without saying a word he went straight up to the grandmother's bed and gobbled her up.

From The Treasury of Fairy Tales, Chichester: Autumn Publishing Ltd. pp.123-126.

CLOZE/86 ENG 1-50

Number	E.word	Good N	Good %	Wrong N	Wrong %	Blank N	Blank %	ID	Range
1.	loved	73	50.0	49	33.6	24	16.4	.41	28
2.	gave	37	25.3	91	62.3	18	12.4	.40	24
3.	velvet	0	0	100	68.5	46	31.5	.0	25
4.	well	15	10.3	91	62.3	40	27.4	.35	18
5.	favourite	17	11.6	97	66.4	32	22.0	.48	17
6.	Riding	146	100	0	0	0	0	0	0
7.	to	130	89.0	8	5.5	8	5.5	.35	4
8.	Hood	146	100	0	0	0	0	.0	0
9.	this	77	52.7	64	43.8	5	3.5	.56	14
10.	Her	94	64.4	49	33.6	3	2.0	.22	4
11.	the	106	72.6	21	14.4	19	13.0	.44	12
12.	anyone/any-body	40	27.4	89	61.0	17	11.6	.61	14
13.	dangerous	9	6.2	111	76.0	26	17.8	.27	17
14.	This	29	19.9	79	54.1	38	26.0	.50	17
15.	delicious	3	2.1	124	85.0	19	13.0	.22	19
16.	than	77	52.7	41	28.1	28	19.2	.56	12
17.	be	57	39.0	34	23.3	55	37.3	.64	20
18.	get	14	9.6	72	49.3	60	41.1	.38	23
19.	a	60	41.1	49	33.6	37	25.3	.52	10
20.	Little	118	80.8	21	14.4	7	4.8	.20	7
21.	said	105	71.9	28	19.2	13	8.9	.42	13
22.	these	70	47.9	71	48.6	5	3.4	.45	14
23.	look	63	43.2	52	35.6	31	21.2	.59	19
24.	hear	30	20.5	71	48.6	46	31.5	.48	19
25.	singing	49	33.6	42	28.7	55	37.7	.57	18

CLOZE/85 ENG 1-50

Number E.word	Good		Wrong		Blank		ID	Range
	N	%	N	%	N	%		
26. you	52	35.6	56	38.4	38	26.0	.40	23
27. you	111	76.0	14	9.6	21	14.4	.43	10
28. realize	3	2.0	119	81.6	24	16.4	.20	14
29. forest	11	7.5	98	67.2	37	25.3	.30	10
30. and	91	62.3	27	18.5	28	19.2	.51	18
31. the	88	60.3	34	23.3	24	16.4	.28	14
32. be	74	50.7	50	34.2	22	15.1	.62	18
33. a	97	66.4	33	22.6	16	11.0	.52	8
34. so	9	6.2	121	82.9	16	10.9	.39	15
35. still	5	3.4	70	47.9	71	48.7	.24	26
36. there	57	39.0	54	37.0	35	24.0	.24	14
37. the	122	83.6	6	4.1	18	12.3	.47	5
38. she	117	80.1	8	5.5	21	14.4	.51	6
39. more	30	20.6	54	37.0	72	49.4	.54	11
40. went	43	29.5	34	23.3	69	47.3	.52	19
41. and	95	65.1	14	9.6	37	25.3	.57	10
42. however	6	4.2	70	47.9	70	47.9	.26	26
43. and	126	86.3	17	11.6	3	2.1	.25	9
44. outside	1	0.7	142	97.3	3	2.0	.0	9
45. I	146	100	0	0	0	0	0	0
46. open	134	91.8	9	6.2	3	2.0	.29	4
47. push	60	41.1	53	36.3	33	22.6	.66	19
48. I	144	98.6	0	0	2	1.4	.11	0
49. up	78	53.4	39	26.7	29	19.9	.29	14
50. the	131	89.7	3	2.1	12	8.2	.33	3